Henry, L., Messer, D. J., Luger-Klein, S. & Crane, L. (2012). Phonological, visual, and semantic coding strategies and children's short-term picture memory span. The Quarterly Journal of Experimental Psychology, 65(10), pp. 2033-2053. doi: 10.1080/17470218.2012.672997

Phonological, visual and semantic coding strategies and children's short-term picture memory span

Henry, L.A.[1], Messer, D.[2], Luger-Klein, S.[1] & Crane, L.[1]

[1] London South Bank University

[2] Open University

Correspondence:

Lucy A. Henry

Department of Psychology

London South Bank University

103 Borough Road

London, UK SE1 0AA

00 44 7771 626936 henrylc@lsbu.ac.uk

Other contact details:

David Messer: (d.j.messer@open.ac.uk):

Scarlett Luger-Klein (scarlettlk@yahoo.com): LSBU address

Laura Crane (cranel3@lsbu.ac.uk): LSBU address

Running Head: Coding strategies in picture memory span

Keywords: picture memory span; phonological, visual and semantic similarity effects; word length effects, memory development; typical children

Acknowledgements:

We would like to thank Willowbrook Primary School, Oval Montessori nursery, St John's Upper Holloway CE Primary School, Clerkenwell Parochial CE Primary School and several other schools and nurseries who very kindly gave their time to help with this project. We would also like to thank Jess Darby, Kathy Filer, Sally Marlow, Gilly Nash and Shelley Revell for their help with recruitment; and Keely Frasca for invaluable help with testing.

Phonological, visual and semantic coding strategies and children's short-term picture memory span

Abstract

Three experiments addressed controversies in the previous literature on the development of phonological and other forms of short-term memory coding in children, using assessments of picture memory span that ruled out potentially confounding effects of verbal input and output. Picture materials were varied in terms of phonological similarity, visual similarity, semantic similarity, and word length. Older children (6/8-year-olds), but not younger children (4/5-year-olds), demonstrated robust and consistent phonological similarity and word length effects, indicating that they were using phonological coding strategies. This confirmed findings initially reported by Conrad (1971), but subsequently questioned by other authors. However, in contrast to some previous research, little evidence was found for a distinct visual coding stage at 4 years, casting doubt on assumptions that this is a developmental stage that consistently precedes phonological coding. There was some evidence for a dual visual and phonological coding stage prior to exclusive use of phonological coding at around 5-6 years. Evidence for semantic similarity effects was limited, suggesting that semantic coding is not a key method by which young children recall lists of pictures.

Children's ability to use verbal mediation for phonological recoding and rehearsal of non-verbal information has long been of interest (Conrad, 1971; Flavell, Beach & Chinsky, 1966; Flavell, Green, Flavell & Grossman, 1997) and represents a major cognitive advance that may be linked to reading development (Palmer, 2000b) and other cognitive skills. For example, Zelazo and colleagues have focussed on the way that non-verbal information and strategies become coded into a phonological/verbal form, which facilitates problem-solving and executive functioning in pre-school children (e.g. Marcovitch & Zelazo, 2009; Zelazo, 2004). In older children, the ability to use verbally coded, explicit information has been an important aspect of the representational redescription model of Karmiloff-Smith (1992) and of subsequent research supporting this model (Messer, Pine & Butler, 2008; Pine & Messer, 2003). Importantly, phonological coding and self-regulatory private speech seem to be related. Al-Namlah, Fernyhough and Meins (2006) argued that both skills underpin the important developmental transition to verbal mediation in the early school years, and Winsler and Naglieri (2003) provided an in-depth account of the development of 'private speech' and problem-solving strategies in school-age children, concluding that children's verbal strategies became internalised with age.

This paper revisits classic research on the phonological coding of non-verbal information, but takes into account important methodological considerations overlooked in previous studies. According to the working memory model (Baddeley, 1986; 2000; 2007; Baddeley & Hitch, 1974), phonological coding occurs within the phonological loop, which consists of two components: the phonological store and the articulatory rehearsal process. For auditorily presented items there is direct access to the phonological store because phonological codes are created by the vocal input itself. Visually presented items such as nameable pictures can only enter the phonological store indirectly after a phonological code has been created; this is achieved by recoding visual material using the articulatory rehearsal process (subvocal articulation). Unless prevented by articulatory suppression, adults use phonological coding to recall visually presented nameable stimuli as evidenced by phonological similarity effects whereby the recall of similar-sounding letters/words is poorer than the recall of dissimilar sounding items (e.g. Baddeley, 1966; Conrad & Hull, 1964). Baddeley (1986, 2007) suggested that phonological similarity effects are attributable to confusion in the phonological store, but there may also be confusion at the recall or reconstruction/redintegration stage (Cowan, Saults, Winterowd & Sherk, 1991; Hasselhorn & Grube, 2003).

Conrad's (1971) classic finding was that children developed 'speech' coding by the mental age of 5-6 years: i.e. they used phonological coding to remember lists of nameable pictures, hence became sensitive to phonological similarity. However, subsequent work by Hulme (1987) contradicted these findings, showing that *all* children of 4, 7 and 10 years showed modest effects of phonological similarity, and that the magnitude of these effects did not change with age. Similarly, Cowan et al. (1991) found effects of phonological similarity in 4-year-olds, questioning whether phonological coding followed the developmental progression put forward by Conrad. Yet close inspection of the methods used in these studies indicates that some form of 'naming' occurred as pictures were presented, either by the experimenter or the child, introducing auditory input and making the results difficult to interpret. To examine whether children *spontaneously* use phonological coding, auditory input should be avoided. Naming pictures at presentation could produce a verbal representation of the items, due to the 'obligatory' phonological coding of heard materials in the phonological store (Baddeley, Lewis & Vallar, 1984).

Halliday, Hitch, Lennon and Pettipher (1990) replicated Conrad's findings using 'silent' presentation, finding phonological similarity effects in 10- but not 5-year-olds. Hitch, Halliday, Schaafstal and Heffernan (1991) went on to underline the importance of auditory input, demonstrating that phonological similarity effects *emerged* in 5-year-olds when either the child or the Experimenter 'labelled' pictures during presentation (see also Ford & Silber, 1994; but note Palmer, 2000b, who argued that preventing labelling by the child made no difference to the findings). Hitch et al. (1991)

argued that overt speech 'primed' the use of phonological representations in young children who would not otherwise use phonological coding with visual items. More recently, Al-Namlah et al. (2006) reported phonological similarity effects in children below the age of 6 years using 'silent' presentation conditions, but like all of the previous studies, they employed *spoken recall*, which makes it necessary for the child to plan and execute speech output. This could affect the use of phonological coding, encouraging the use of a speech code in preparation for a verbal response. The current studies, therefore, always required a non-verbal response.

Span procedures determine the longest lists a child can recall in serial order, whereas fixed list length procedures involve administering groups of same-aged children the same list lengths chosen to be at or around their ability level (e.g. Hulme, 1987, used lists of four items for four-year-olds, five items for seven-year-olds and six items for 10-year-olds). One problem with fixed list length procedures is that task difficulty is not the same for individual children: some receive "supra span" lists beyond their span level; some receive lists at span level; and some receive lists below span level. The mechanisms underlying span and supra-span performance cannot be assumed to be identical. For example, strategic approaches may differ when span capacity is overloaded: McGilly and Siegler (1989) found more evidence for rehearsal when *shorter* lists were presented to young children. Two further methodological issues are as follows: Hitch et al. (1991) noted that evidence for 'similarity' effects is enhanced by blocking lists of each type together (i.e. not alternating similar and dissimilar picture types across trials), to increase proactive interference. Jarrold, Cocksey and Dockerill (2008) recommended full recall over probed recall methods, as the latter produce high recency on later serial positions, increasing ceiling effects and reducing task sensitivity to detect phonological similarity effects.

The current studies revisited the development of phonological similarity effects in children between the ages of 4 and 8 years, using picture span procedures that ensured that evidence of phonological coding reflected deliberate strategy choices. Purely visual presentation of pictures was achieved by avoiding naming or labelling from either the experimenter or the child. Non-verbal recall via pointing to pictures in serial order removed any requirement for verbal output. Span procedures with full recall ensured that all children were assessed sensitively at their own individual span level (Experiments 1 and 2) or at a level just above span (Experiment 3) and picture type was always blocked, rather than alternated, to maximise the evidence for similarity effects.

Word length effects, whereby items with long names are less well recalled than items with short names (Baddeley, Thomson & Buchanan, 1975; Hulme, Thomson, Muir & Lawrence, 1984), may also indicate phonological coding in picture span tasks (Gathercole & Hitch, 1993; Henry & Millar, 1993). Current theories assume that the phonological characteristics of longer words are more difficult/complex to deal with than short words (e.g. spoken duration or differences in phonological complexity – see Hulme, Neath, Stuart, Shostak, Surprenant & Brown, 2006; Lewandowsky & Oberauer, 2008; Mueller, Seymour, Kieras & Meyer, 2003; Romani, McAlpine, Olson, Tsouknida & Martin, 2005). Word length effects in picture span tasks emerge at 7-9 years (Halliday et al., 1990; Henry, Turner, Smith & Leather, 2000; Hitch, Halliday, Dodd & Littler, 1989; Hitch et al., 1991), although there is debate about the precise cognitive processes responsible for this development in relation to verbal rehearsal and verbal output (Cowan, Day, Saults, Keller, Johnson & Flores, 1992; Henry, 1991; Henry et al., 2000; Yuzawa, 2001). Others doubt that word length effects require articulatory processes (e.g. Hulme et al., 2006; Romani et al., 2005), but for current purposes, word length effects were examined as an additional indication of phonological coding.

Picture span methods also constitute a useful means of identifying visual coding, which is assumed to be present if an individual has better memory for visually dissimilar than visually similar items. This is attributed to interference caused in the memory trace when trying to store and retrieve items with similar visual characteristics (e.g. Frick, 1988a; 1988b; Hitch, Halliday, Schaafstal, & Schraagen, 1988;

Logie, Della Sala, Wynn & Baddeley, 2000). Visual coding has often been reported in samples of five-year-olds but not in older children (Brown, 1977; Hayes & Schulze, 1977; Hitch et al., 1988; Hitch, Woodin & Baker, 1989; Hitch et al., 1991; Longoni & Scalisi, 1994; Palmer, 2000a). Palmer (2000a) outlined a stage-like development in strategy use, noting that 3-4-year-old children used neither visual nor phonological strategies, and may rely on semantic representations or "automatic activation of representations in long-term memory" (although this was not directly tested). Palmer argued that by 5 years, children utilised visual coding, and that phonological coding emerged after the age of 5 or 6 years. Between the ages of 6 and 8 years children used both visual and phonological coding strategies (dual coding), before predominantly phonological coding strategies emerge by 10 years (Hitch et al., 1988; Hitch et al., 1989; 1991). However, such conclusions about phonological and other forms of coding may be weakened by methodological confounds, so the present studies were designed to shed further light on this stage model.

Finally, Palmer's speculation that semantic coding may be an early developing strategy could be incorrect; most models of the lexicon suggest connections between semantic and phonological representations (e.g. Levelt, Roelofs & Meyer, 1999; Messer & Dockrell, 2006), so semantic coding could be a later developing strategy in children already using phonological coding. Adults do use semantic coding: Baddeley (1966) found that semantic similarity had a small but negative effect on recall. Yet Purser and Jarrold (2010) reported that typical 4-6-year-old children used both phonological and semantic storage in a 'matching span' task. Given this lack of clarity, the current studies tested two opposing views of the use of semantic codes: (1) They are used by developmentally less mature children in the stage prior to phonological coding (Palmer, 2000a); or (2) semantic coding is a more mature strategy, closely tied to the development of phonological coding because of automatic links between semantic and phonological information in long-term memory (e.g. Huttenlocher & Lui, 1979). We defined semantic similarity in terms of concrete, highly imageable nouns that shared a conceptual category (body parts).

In summary, this paper addresses a number of controversies in the previous literature on phonological and other forms of coding in a series of methodologically sound and carefully controlled experiments. Potential confounds in interpreting developmental changes in memory coding in picture memory span tasks were avoided by eliminating auditory input and removing verbal output. Individual differences with respect to strategic development *within age groups* (Palmer, 2000b) could also make generalisations based on age problematic, so traditional 'age differences' analyses were followed by cluster analyses to identify groups of children with similar abilities and explore the memory coding strategies used in each resulting cluster. Potentially reduced task sensitivity due to low span levels was tackled in Experiment 3 with a new procedure to improve task sensitivity and reliability. 'Span' assessments were retained, but children were pre-assessed for *baseline* picture memory span level and then tested on 'experimental' picture types using span level "plus 1" (span+1) length lists to ensure that all experimental lists were equally demanding across participants, but slightly longer than they would have been using a typical span procedure. Some span+1 assessments were repeated one week later to test reliability.

It was expected that younger children (4-5 years) would use visual coding strategies (visual similarity effects), and older children (6-8 years) would use dual coding, namely, both visual and phonological strategies (visual similarity, phonological similarity and word length effects). We tentatively predicted that semantic similarity effects would be more likely to be found in our older group because of automatic links between semantic and phonological information in long-term memory.

Experiment 1

Method

Design

Picture memory span for four types of pictures (phonologically similar, visually similar, long names, semantically similar) and a control set of pictures was assessed in two age groups of children (4/5, 6/8 years). Age level was a between participants factor, and picture type was a within participants factor.

Participants

The sample comprised 18 children of 4-5 years (mean age 4 years 11 months, SD 10 months, 8 boys) and 27 children of 6-8 years (mean age 7 years 5 months, SD 8 months, 12 boys). Children attended mainstream schools and nurseries in London and South East England and had no special needs identified. In all three experiments, written informed consent from both parents and children (using a specially worded consent form) was obtained prior to testing and ethical approval was given by the Research Ethics Committee of London South Bank University.

Apparatus / Materials

Each set of pictures comprised nine 10cm x 15cm white cards with black on white line drawings of familiar, highly imageable objects. All pictures were hand drawn in pencil and black ink on white paper using the same straightforward style of depiction (for examples please see Henry, 2008), then scanned as images for reproduction. Copies of all picture materials are available on application from the first author.

Control items were neither phonologically, semantically nor visually similar (cake, chair, shoe, bus, leaf, frog, ring, clown, drum).

Phonologically similar items (can, lamp, hat, van, pan, ant, cat, bat, fan) shared the same vowel, [æ], regarded as the most important factor for phonological similarity (Nimmo & Roodenrys, 2004).

Visually similar items (bed, fish, nail, key, sock, tie, pen, knife, brush) were similar in terms of shape (all were depicted as very long thin objects drawn at the same 45° angle).

Long-named items (bicycle, teddy bear, umbrella, television, elephant, butterfly, ladybird, telephone, banana) were three/four syllables in spoken length.

Semantically similar items (thumb, neck, hair, leg, arm, eye, teeth, ear, lips) all came from the same conceptual category.

All objects in the long-named stimulus set had three/four-syllable names, whereas the items for all other conditions had one-syllable names. Except for the phonological similarity condition, all sets contained objects with phonologically dissimilar names. Apart from the visual similarity condition, objects in each set were visually dissimilar. Apart from the semantically similar items, objects from each set represented a diverse range of categories. Item sets were matched as closely as possible in terms of mean age of acquisition of object names, imageability, frequency, familiarity, and name agreement (Morrison, Chappell & Ellis, 1997). Note that ratings were not available for two items in the phonologically similar set (fan, can), two items in the semantically similar set (neck, teeth), and one item in the long name list (teddy bear); but due to constraints in selecting appropriate materials, they were included. Table 1 shows the mean ratings (and ranges) available for each set. It proved difficult to obtain semantically similar body part names that were closely matched to the other sets on age of acquisition and familiarity; however, for this reason, the items 'neck' and 'teeth' were included as they were regarded as slightly less familiar. Similarly, it proved difficult to select longer items with the highest levels of name agreement (i.e. 1), but all items chosen were as close to this as possible, given other constraints.

Table 1 about here

There were 25 21cm x 30cm response boards, with five for each set of picture stimuli. The response boards all differed, with each containing a random arrangement of the nine objects from that set in a 3 x 3 array. The response boards were changed after each trial; this prevented participants from learning the spatial locations of the items during each picture memory span task.

Procedure

Children were tested in a classroom at school or a room in their family home. The testing took place in one long session of approximately 30-40 minutes or in two shorter sessions of approximately 15-20 minutes, depending on the child's and/or the parents' choice. Participants were asked to look at a series of picture cards, presented one at a time, and to point to the same pictures in the same order on the response board. The importance of serial order report has been documented by a number of researchers (Avons, 1998; Avons & Mason, 1999; Avons, Ward & Melling, 2004).

Each picture card was presented for 1.5 seconds directly in front of the child, and then removed from view. In previous research, pictures have typically been presented in a fixed horizontal row, turning each card face down in turn (Hitch et al., 1988; Hitch et al., 1989; Hitch, Halliday et al., 1989; 1991; Longoni & Scalisi, 1994). This provides the potential for spatial cues to assist performance, so cards were always presented in the same spatial location (Henry, 2008) to make the task comparable to verbal span tasks. The experiment began with a set of four practice pictures and span trials of one and two items; once the child understood the task he/she moved on to experimental trials.

The order of experimental picture memory span tasks was counterbalanced according to a Latin square. Before each span task, the child was presented with the nine cards of the set, and objects were named for them by the experimenter to ensure that children were familiar with all presented items and their names; none of the children appeared unfamiliar with any of the items in the pictures. Children did not name the pictures to reduce priming for phonological strategies.

To determine the maximum number of items a child could recall on each span task, up to six lists were administered at any given list length. For the youngest participants, experimental trials started with a list length of one item, but lists of one were omitted for the oldest children or those whose spans were comfortably higher than two items on earlier sets. For each list length, if the child passed fewer than four trials, they were deemed not to have passed that span level, and testing ceased. If the child passed at least four of six trials, list length was increased by one item and another set of trials started. Each time a higher list length was introduced, children were informed about how many pictures would be shown to them next, and they were reminded each time to point to them in the correct order. Children were instructed not to name items out loud during presentation of pictures in order to avoid adding overt verbal input (Hitch et al., 1991). If they were observed to do so, they were reminded again. Most children followed this instruction from the beginning, and no child needed more than two reminders.

Results

Mean memory spans for each picture type at each age level are given in the top part of Table 2. A mixed analysis of variance (ANOVA) including the repeated factor 'picture type' (five levels) and the between participants factor 'age group' (4/5 years, 6/8 years) revealed main effects of picture type, $F(4,172) = 11.70$, $p < .001$, partial $\eta^2 = .21$, and age group, $F(1,43) = 35.99$, $p < .001$, partial $\eta^2 = .46$; and a significant picture type by age group interaction, $F(4,172) = 3.95$, $p < .01$, partial $\eta^2 = .08$.

Given the significant interaction and our predictions, one-way ANOVAs were used to explore the effects of picture type in each age group, followed by planned contrasts to test between ‘control’ item performance and performance on all other picture types. In this and all subsequent analyses of picture type effects using planned contrasts/comparisons, significance levels were adjusted to reflect the four comparisons carried out: i.e. 0.05/4 = $p < .0125$. For the 4/5-year-old children, there was no effect of picture type ($F = 1.64$). For the 6/8-year-old children, there was a significant effect of picture type, $F(4,104) = 15.44$, $p < .001$, partial $\eta^2 = .37$. Planned contrasts showed significant reductions in recall for phonologically similar items, $F(1,26) = 43.79$, $p < .001$, partial $\eta^2 = .63$, a 'phonological similarity effect’ (PSE); and for long named-items, $F(1,26) = 17.33$, $p < .001$, partial $\eta^2 = .40$, a 'word length effect' (WLE). Visually similar and semantically similar item recall did not differ from that for control items.

Table 2 about here

It is possible that variability in strategy use among younger children concealed the expected visual similarity effects. Therefore, a hierarchical cluster analysis was carried out using performance across the five sets of pictures to assess whether distinct clusters could be identified and, if so, the nature of any group differences between them. The cluster analysis was carried out in SPSS16. A between group linkage method was used that involved squared Euclidean distance. The cluster analysis was based on five standardised *z*-scores derived from each of the five span scores. An initial analysis was run with a limit of four clusters as this corresponded to the 'stages' that have been identified in the development of picture span strategies. Only six children were identified in the smallest cluster, this together with an inspection of the dendrogram, resulted in the decision to base our analyses on a three cluster solution. Inspection of the profiles of the three cluster solution revealed differences in the proportion of correct responses across the three clusters, with the lowest performing cluster having a flat profile (data illustrated in Figure 1).

The mean ages of children in the clusters matched their performance (mean age for the cluster with the highest performance was 93.7 months, SD 8.5, N = 18; mean age for the cluster with intermediate performance was 71.3 months, SD 8.4, N = 14; and mean age for the cluster with the lowest performance was 60.6 months SD 12.0, N = 13). These age differences were significant ($F(2,42) = 49.03$, $p < .001$), and Bonferroni post hoc tests indicated that all comparisons involved significant differences ($p < .001$, except for the comparison of the intermediate and low clusters where $p < .05$).

A 3 (cluster) x 5 (picture type) mixed ANOVA on the picture memory span scores revealed main effects of cluster, $F(2,42) = 148.11$, $p < .001$, partial $\eta^2 = .88$, and picture type, $F(4,168) = 13.46$, $p < .001$, partial $\eta^2 = .24$, and a significant interaction, $F(8,168) = 3.23$, $p < .002$, partial $\eta^2 = .13$. To investigate whether there were differences in performance according to picture type, one-way repeated measures ANOVAs were carried out on the span data from each cluster. Figure 1 illustrates the findings.

For the highest performing cluster there was a significant effect of picture type, $F(4,68) = 15.00$, $p < .001$, partial $\eta^2 = .47$. Planned comparisons indicated significant reductions in recall for phonologically similar items ($p < .001$, PSE) and long named-items ($p < .01$, WLE). For the intermediate cluster, there was a significant effect of picture type, $F(4,48) = 3.38$, $p < .05$, partial $\eta^2 = .22$, and planned comparisons showed significant reductions in performance for phonologically similar items ($p < .01$, PSE). In the lowest performing cluster there were no significant differences in performance across picture types.

Figure 1 about here

Summary. Analyses of age differences in picture span scores indicated that 4/5-year-olds did not show visual similarity, phonological similarity, semantic similarity or word length effects; whereas 6-8-year-olds showed phonological similarity and word length effects. These findings suggested that younger children used no particular form of coding and older children used phonological coding (and possibly verbal rehearsal strategies). When the results were assessed in relation to clusters of performance rather than age, a slightly more nuanced picture emerged. The highest performing children showed phonological similarity and word length effects consistent with phonological coding (and possibly verbal rehearsal); the middle performing children showed phonological similarity effects consistent with phonological coding; and the lowest performing children showed no effects, consistent with a lack of any particular coding strategy.

Experiment 2

Experiment 1 raised several issues concerning the use of different forms of coding in picture span tasks. The lack of evidence for visual coding in the youngest groups was similar to some previous reports (Henry, 2008; Hitch et al., 1989b), but dissimilar to other positive findings of visual similarity effects (Hitch et al., 1988; 1989a; 1991; Palmer, 2000a). The differing findings could be accounted for by the more stringent methodology employed here, but replication with a slightly larger sample of 4/5-year-old children was deemed desirable.

Method

Participants

Fifty-nine children were recruited: 44 from two primary schools in North London, and a further 15 children via social contacts of the experimenter. Three children were excluded from the study because they did not spontaneously use English as their primary language of communication. The final sample comprised 26 children aged 4-5 years (mean age 5 years 0 months, SD 6 months, 12 boys) and 30 children aged 6-7 years (mean age 6 years 9 months, SD 5 months, 14 boys).

Apparatus/materials/procedure

These were identical to Experiment 1 with two exceptions. First, three trials per list length were administered instead of up to six in order to reduce testing time. Span was scored as the total number of trials correct, as this measure may be more sensitive than span (Ferguson, Bowey & Tilley, 2002). Secondly, to minimise the effects of physical enactment strategies, which had occasionally been noted in Experiment 1, children were told “this is not a game where we use our body or our hands”, and reminders were given as necessary. Children were also given some additional tests not reported here.

Results

The means and standard deviations of the five picture span measures (total number of trials correct) for children in each age group are presented in the middle part of Table 2. A mixed two-way ANOVA with 'age group' as the between-participants factor and 'picture type' as the within-participants factor was carried out. Age group had a significant effect on memory scores, $F(1, 54) = 20.49$, $p < .001$, partial $\eta^2 = .28$, indicating that older children had higher scores than younger ones. The effect of picture type was significant, $F(4, 216) = 10.21$, $p < .001$, partial $\eta^2 = .16$; as was the interaction between age group and picture type, $F(4, 216) = 2.92$, $p < .05$, partial $\eta^2 = .05$ (Greenhouse-Geisser corrections for violation of sphericity, $p = .04$, gave the same results).

The interaction between age and picture type was, again, of interest, so separate one-way ANOVAs were carried out for the two age groups. For 4/5-year-olds, picture type had no significant effect on memory span (F=1.90). For 6/7-year-olds, picture type had a significant effect on memory span, $F(4, 116) = 9.96$, $p < .001$, partial $\eta^2 = .26$. Planned contrasts revealed memory span reductions for phonologically similar items (PSE), $F(1, 29) = 26.14$, $p < .001$, partial $\eta^2 = .47$; visually similar items (VSE), $F(1, 29) = 13.20$, $p < .001$, partial $\eta^2 = .31$; and long-named items (WLE), $F(1, 29) = 18.17$, $p < .001$, partial $\eta^2 = .39$.

The same cluster analysis was run as for Experiment 1. The four cluster solution produced two clusters that contained 7 and 8 participants respectively, and consequently, a three cluster solution was selected as this was supported by inspection of the dendrogram to give larger numbers in fewer groupings. The performance of children in the three clusters is shown in Figure 2. An ANOVA revealed there was a significant difference in the ages of the children in the three clusters, $F(2,52) = 6.20$, $p < .004$. Bonferroni post-hoc comparisons indicated that the highest performing cluster (N= 8, mean age 81.3 months, SD 4.6) was significantly older than the other two ($p < .004$; intermediate cluster 1, N

= 24, mean age 72.7 months, SD = 11.6; lowest performing cluster, N = 23, mean age 66.1 months, SD 11.4).

A 3 (cluster) x 5 (picture type) mixed ANOVA on the picture span scores showed main effects of cluster, $F(2, 52) = 127.43$, $p < .001$, partial $\eta^2 = .83$, and picture type, $F(4, 208) = 17.36$, $p < .001$, partial $\eta^2 = .25$; and a significant interaction, $F(8, 208) = 7.42$, $p < .001$, partial $\eta^2 = .17$. Post-hoc Bonferonni tests showed all the differences between groups to be significant at $p < .001$.

One-way ANOVAs were then carried out to assess the effects of picture type in each cluster. For the highest performing cluster there was a significant effect of picture type, $F(4,28) = 10.221$, $p < .001$, partial $\eta^2 = .59$; and planned comparisons revealed significant performance reductions on phonological ($p < .01$, PSE), visual ($p < .01$, VSE), and long-named ($p < .01$, WLE) pictures. For the cluster with intermediate performance there was an effect of picture type, $F(4,92) = 5.51$, $p < .01$, partial $\eta^2 = .19$; and planned comparisons revealed significant performance reductions on phonological ($p < .01$, PSE), long-named ($p < .01$, WLE), and semantic (p < .01, SSE) pictures (this approached significance for visual pictures, $p = .016$). For the lowest performing cluster there was a modest overall effect of picture type, $F(4,88) = 3.40$, $p < .05$, partial $\eta^2 = .13$; however, planned comparisons did not identify any significant differences.

Figure 2 about here

Summary. In Experiment 2, the least mature children, by age or cluster, showed no evidence for any form of coding; whereas more mature children showed clear evidence for phonological coding (PSE, WLE). There was some evidence for visual and semantic coding, but only in more mature children.

Experiment 3

Experiment 3 explored an alternative method for assessing picture span performance that reduced insensitivity to task differences at low levels of performance, yet still avoided the drawback of fixed list lengths (too easy or difficult for many participants). "Baseline" span was assessed for every participant and subsequently 10 trials for each ‘experimental’ picture type were administered at a level of baseline span plus one (span+1). This equated difficulty levels for all participants whilst delivering slightly longer and ‘more confusable’ lists for lower performing children. We also asked children how they had remembered the pictures to try to detect self-reported phonological coding. Finally, the reliability of control and phonologically similar span scores was calculated from a retest one week later.

Method

Participants

The sample consisted of 31 children of 4-5 years (mean age 5 years 4 months, SD 6 months, 19 boys) and 35 children of 6-8 years (mean age 7 years 3 months, SD 9 months, 17 boys). The children attended mainstream schools and nurseries in Greater London and had no special needs identified.

Apparatus/materials/procedure

These were the same as previous experiments with the following exceptions. First, all children were assessed for baseline picture memory span using an alternative set of ‘control’ pictures with the same characteristics as the other picture sets (bell, clock, tree, dog, comb, moon, bread, glass, shirt; see Table 1). There were six trials per list length and ‘baseline’ span scores represented the longest list at which the child obtained at least four trials out of six entirely correct (age 4/5 years, mean span = 1.35, range 1-2; age 6/8 years, mean span = 2.6, range 2-4). All subsequent picture memory span tests were administered at a list length of baseline span+1 (if a child's picture memory span was two, all subsequent tests were administered at a list length of three items). Ten trials for each experimental picture type in counterbalanced order were administered over two testing sessions. Scores reflected the number of pictures recalled in the correct order (if one picture was left out, the remainder were counted as incorrect; if the order of two pictures became reversed, subsequent and prior pictures could be counted as correct) and were expressed as proportions correct for each picture type. Approximately one week later, children were retested for their baseline span+1 performance on control and phonological pictures to examine the reliability of these measures. At the end of this session children were also asked how they had remembered the pictures using a series of questions (see Appendix). Children were classified as self-reporting phonological coding if they reported that they 'said the names of the pictures inside their head'.

Results

The lower part of Table 2 includes the *proportion correct* scores for each picture type. An age group (4/5-year-olds, 6/8-year-olds) by picture type (control, phonological, visual, long, semantic) mixed ANOVA revealed significant effects of picture type, $F(4, 256) = 3.56$, $p < .01$, partial $\eta^2 = .053$; and a significant interaction between picture type and age, $F(4, 256) = 3.60$, $p < .01$, partial $\eta^2 = .053$. The effect of age group was not significant, suggesting that our new method had succeeded in equating task difficulty, a necessary precondition for equivalent task sensitivity across age (although younger children still obtained marginally lower scores, $F(1, 64) = 3.1$, $p < .10$, partial $\eta^2 = .046$).

Given the predictions, one-way ANOVAs were used to explore the effects of picture type in each age group. There were no significant effects of picture type for 4/5-year-olds ($F < 1$). Picture type was, however, significant for 6/8-year-olds, $F(4, 136) = 6.86$, $p < .001$, partial $\eta^2 = .17$. Planned contrasts indicated 6/8-year-olds showed reduced performance in the phonological condition, $F(1, 34) = 8.58$, $p < .01$, partial $\eta^2 = .20$; and in the long-named condition, $F(1, 34) = 10.30$, $p < .01$, partial $\eta^2 = .23$ (PSE and WLE).

The same cluster analysis as used in previous experiments was carried on the data from Experiment 3. Both the four and three cluster solutions identified groups with only 3 members, so a two cluster solution was used for analysis (see Figure 3). A 2 (cluster) x 5 (picture type) mixed ANOVA showed main effects of cluster, $F(1, 64) = 256.29$, $p < .001$, partial $\eta^2 = .80$, and picture type, $F(4,256) = 4.29$, $p < .01$, partial $\eta^2 = .06$, and a significant interaction between the two, $F(4,256) = 4.03$, $p < .01$, partial $\eta^2 = .06$.

One-way ANOVAs explored picture type differences for each cluster. For the higher performing cluster, picture type was significant, $F(4,129) = 6.43$, $p < .001$, partial $\eta^2 = .18$; and planned comparisons revealed significant performance reductions on phonological (PSE, $p < .001$) and long - named pictures (WLE, $p < .01$). For the cluster with lower performance there was no overall effect of picture type ($F < 1$).

Figure 3 about here

Self-reported phonological coding. Thirty children reported phonological coding (mean age 84.4 months, SD=12.3 months, range 65-101 months) and 36 did not (mean age 69.6 months, SD=12.3, range 53-105 months). Age differences between these groups were significant, $t(64) = 5.12$, $p < .001$, but the overlap in age ranges was considerable. This indicated that self-reported phonological coding was not perfectly related to age (point biserial correlation between age and phonological coding was in the moderate range, $r = .54$).

A comparison was made of picture recall in the two groups. A phonological coding group (yes, no) by picture type (control, phonological, visual, long, semantic) mixed ANOVA revealed a significant effect of coding group $F(1, 64) = 4.50$, $p < .05$, partial $\eta^2 = .066$ (phonological coders recalled more pictures); picture type, $F(4, 256) = 4.34$, $p < .01$, partial $\eta^2 = .063$; and a significant interaction between picture type and coding group, $F(4, 256) = 4.00$, $p < .05$, partial $\eta^2 = .05$.

One-way ANOVAs on the two self-reported coding groups indicated that self-reported non-phonological coders showed no significant effects of picture type (*F*=1.16). Phonological-coders showed significant effects of picture type, $F(4, 116) = 5.88$, $p < .001$, partial $\eta^2 = .169$, and planned contrasts revealed significant performance reductions on phonologically similar, $F(1, 29) = 8.41$, $p < .01$, partial $\eta^2 = .225$ (PSE), and long-named pictures, $F(1, 29) = 10.45$, $p < .01$, partial $\eta^2 = .265$ (WLE).

Re-test data. This was available for phonological and control picture types. A three-way mixed ANOVA including the factors of age (4/5, 6/8-year-olds), picture type (control, phonological), and test session (time one, retest) was carried out on the proportion correct scores. There was a significant effect of picture type, $F(1, 63) = 9.78$, $p < .01$, partial $\eta^2 = .13$. Inspection of the means indicated that performance was higher for both age groups combined on control items (time one mean = .49, SD = .25; retest mean = .51, SD = .26) than phonological items (time one mean = .43, SD = .21; retest mean = .46, SD = .22). There were no effects of age or test session and no interactions. The correlation between performance in the control condition at time one and retest was moderately high ($r = .72$); as was the correlation between performance in the phonological condition at time one and retest ($r = .72$). These data suggest that proportion correct scores on the span+1 task were reliable.

Finally, some studies have reported data on *difference* scores between control and phonological performance in individual children as an indication of phonological coding. However, such methods may increase error variance, because the difference score involves the error variance from two variables. In fact, the correlation between difference scores at time one and retest was low and non-significant ($r = .17$), suggesting that such difference measures are unreliable.

Additional Analysis Combining Span Scores for Experiments 1 and 2

It is possible that subtle differences in strategy use are masked by combining groups that differ in developmental level or age (Henry, 2008). In order to check whether using combined age groups concealed picture type effects that might emerge in each age group separately, we conducted an additional analysis of the span scores from Experiments 1 and 2 combined, increasing our power to look at developmental trends in four separate age groups (24 4-year-olds, 20 5-year-olds, 29 6-year-olds, 28 7/8-year-olds). See Figure 4 for an illustration of this data. Note that span scores were based on six trials per list length in Experiment 1 and three trials per list length in Experiment 2, so the results should be treated with caution.

An age (4, 5, 6, 7/8 years) by picture type (5 levels) ANOVA revealed significant main effects (with Greenhouse-Geisser corrections for violation of sphericity where relevant) of picture type, $F(3.7, 359.9) = 16.22$, $p < .001$, partial $\eta^2 = .14$, and age group, $F(3, 97) = 25.36$, $p < .001$, partial $\eta^2 = .44$. There was also a significant picture type by age group interaction, $F(11.13, 359.9) = 2.96$, $p < .01$, partial $\eta^2 = .08$.

Subsequent one-way ANOVAs for each age group showed an absence of picture type effects in 4-year-olds ($F = 1.64$), but such effects were significant in all other age groups [5-year-olds: $F(4, 76) = 4.05$, $p < .01$, partial $\eta^2 = .17$; 6-year-olds: $F(4, 112) = 7.67$, $p < .001$, partial $\eta^2 = .22$; 7/8-year-olds: $F(4, 108) = 10.75$, $p < .001$, partial $\eta^2 = .29$]. Planned contrasts indicated 5-year-olds showed a PSE, $F(1, 19) = 12.79$, $p < .01$, partial $\eta^2 = .40$; a WLE, $F(1, 19) = 8.64$, $p < .01$, partial $\eta^2 = .31$, and a marginally significant VSE, $F(1, 19) = 5.15$, $p = .035$, partial $\eta^2 = .21$. Six-year-olds showed a PSE, $F(1, 28) = 34.92$, $p < .001$, partial $\eta^2 = .56$; a WLE, $F(1, 28) = 11.67$, $p < .01$, partial $\eta^2 = .29$; a marginally significant VSE, $F(1, 28) = 4.28$, $p = .048$, partial $\eta^2 = .13$; and a marginally significant SSE, $F(1, 28) = 4.48$, $p = .043$, partial $\eta^2 = .14$. Seven/eight-year-olds showed a PSE, $F(1, 27) = 13.93$, $p < .01$, partial $\eta^2 = .34$; and a

WLE, $F(1, 27) = 24.65$, $p < .001$, partial $\eta^2 = .48$ (Fs for VSE and SSE were virtually zero: .16 and .19 respectively). Findings are discussed below.

Figure 4 about here

General Discussion

The present experiments addressed controversies in the previous developmental literature on phonological and other forms of coding, using assessments of picture memory span that controlled for potentially confounding effects of verbal input in children of 4-8 years. Clear and robust support was found for Conrad's (1971) contention that verbal mediation in the form of phonological coding develops gradually in typical children from age 5-6 years. Children in the current samples who were older (6-8 years), more mature in terms of performance levels (more mature "clusters"), and who 'self-reported' the use of phonological coding showed significant phonological similarity and word length effects. By contrast, children who were younger (4-5 years), less mature in terms of performance levels, and who did not self-report the use of phonological coding showed no evidence of strategy use (no negative effects of different picture types). These findings were consistent across three experiments that used different methods of assessing 'span' performance ('traditional' span measures, measures of total trials correct on span tasks, percentages correct on span level+1 lists) and different ways of dividing our groups (based on age, or level of performance). In a combined age-difference analysis of span scores from Experiments 1 and 2, the findings were slightly more nuanced: 4-year-olds showed no evidence of strategy use, but 5- year-olds resembled older children, with evidence emerging for phonological coding.

Theoretically, these findings add to the evidence of a "profound change in the use of verbal mediation" (Al-Namlah et al., 2006, p. 117) over the age of 5 years in typically developing children. The current study has examined phonological coding, but evidence is accumulating that the use of private speech for behavioural regulation and problem-solving is a closely related process (Al-Namlah et al., 2006; Winsler & Naglieri, 2003). Further investigation of both of these important skills will contribute to our understanding of verbal mediation, perhaps confirming that it emerges from the gradual internalisation of social speech as proposed by Vygotsky (1934/1986).

Experiment 3 suggested that the average age of *self-reported* phonological coding was exactly 7 years. Flavell et al. (1997) found, similarly, that 6- to 7-year-olds have a reasonable knowledge and awareness of inner speech, possibly because experience of reading, writing and arithmetic in the early school years requires such skills. Children who said they used phonological coding were able to recall longer lists of items, suggesting that use of this strategy was positively related to the level of recall. Note that early strategy use is not always related to improvements in recall as illustrated by work on ‘utilisation deficiencies’, whereby children use a strategy but gain no resultant benefit (e.g. Miller, 1994). In Experiment 1, there was some evidence of an ‘intermediate’ strategy stage of phonological coding on its own (phonological similarity effects) followed by a later stage of phonological coding plus verbal rehearsal (phonological similarity and word length effects) in the cluster analysis. However, given the difficulties of interpreting word length effects as evidence for verbal rehearsal (Cowan et al., 1992; Henry, 1991; Henry et al., 2000; Yuzawa, 2001), the cautious conclusion is that both of these effects indicated phonological coding. In all other relevant analyses, phonological similarity and word length effects coincided in the older and more mature children.

The findings provide evidence that picture span measures do index the development of phonological coding in children and should be of value in future investigations exploring links between phonological coding and related academic skills such as reading, spelling, writing and arithmetic (e.g. Palmer, 2000b). Picture span methods are also useful tools to investigate the development of phonological coding in atypical populations of children who may have weak language skills and limited abilities to introspect on inner speech use (Henry, 2008; Whitehouse, Maybery & Durkin, 2006; Williams, Happé

& Jarrold, 2008; Williams & Jarrold, 2010). However, our results emphasise that 'difference' scores between span performance on control versus phonologically similar items should be used with caution. In Experiment 3, such measures were not found to be reliable when children were retested after one week.

Experiments 1 to 3 did not offer clear support for the *transition* from visual to phonological coding between four and six years described in previous research (e.g. Palmer, 2000a). Visual similarity effects were not detected in the youngest children, and these children showed no evidence of alternative strategy use (phonological or semantic coding). Some previous research employing span-based assessments has indicated a similar lack of strategy use in young children (Henry, 2008; Hitch et al., 1989), but studies employing fixed list length picture span assessments with verbal recall tend to reveal visual similarity effects in 4/5-year-olds (Hitch et al., 1988; Hitch et al., 1989; 1991; Palmer, 2000a). We will return to this point shortly in discussing the combined analysis of span scores for Experiments 1 and 2.

There was some evidence for the presence of *dual* phonological and visual coding, an interim stage proposed by Palmer (2000a), amongst the more mature children in Experiment 2. Given the variability in the identification of visual and dual coding here and in previous research (Henry, 2008; Palmer, 2000a), we conducted an additional analysis of the span scores from Experiments 1 and 2 combined, increasing the power to examine strategy use in four separate age groups (4-year-olds, 5-year-olds, 6-year-olds, 7/8-year-olds). Although we still found no significant effects of picture type in 4-year-olds, 5-year-olds showed evidence of using phonological coding (significant PSE and WLE) and visual coding (marginally significant VSE). This evidence for dual visual and phonological coding was repeated in the 6-year-olds (significant PSE and WLE, marginally significant VSE). By 7/8 years, visual similarity effects had disappeared, and children showed only phonological coding (PSE, WLE), concurring with our previous findings for older groups and the more mature clusters.

These analyses should be treated with caution because we have combined data over two experiments that did not measure span in exactly the same way. However, the findings highlight several points. First, phonological coding may emerge earlier than our separate experimental results had indicated (at 5 years rather than 6 years). Combining groups of 4- and 5-year-olds may mask differential effects at each age level, an important point for future research hoping to track developmental changes in phonological and visual coding strategies. Second, the findings provided some evidence for a dual phonological and visual coding stage at 5 to 6 years, although the VSE remained marginally significant. It is not clear why some samples of children fail to show consistent and reliable effects of visual similarity, but differences in teaching methods, perhaps in relation to visual versus phonological reading strategies, could affect the use of pure phonological versus dual coding. Finally, the findings give reassurance that picture type effects with span tasks can be found at low levels of recall. Five-year-olds recalled on average 1.6 to 2.3 pictures, yet still showed significant effects of picture type in line with previous reports of similarity effects in children with performance levels of 2 or less (e.g. Hitch et al., 1988; Hitch et al., 1989). Even so, some children still had spans of just one picture. For this reason the span+1 methodology adopted in Experiment 3 offers the advantage of never testing children on lists of less than two pictures.

Although the current findings suggest that visual coding strategies were not a dominant feature of span performance in 4-year-olds, it remains possible that the current experiments were unable to detect them. Ideally our findings would be confirmed using even larger samples of 4-year-olds and greater numbers of trials at the span+1 level. Such data is important to assess properly existing models of short-term memory development based on a hypothesised transition from visual to dual and then exclusive phonological coding (Hitch et al., 1988; Hitch et al., 1989; 1991; Palmer, 2000). Future research could also compare non-verbal with verbal recall methods, to test whether this methodological factor affects the development of dual and visual memory coding strategies.

Finally, the present results offered little evidence for children's use of semantic coding. Younger children were *not* found to show semantic similarity effects in any of the analyses across the three experiments, providing no direct evidence for a developmentally early 'semantic coding' phase (Palmer, 2000a). There was limited evidence for semantic similarity effects in the intermediate cluster of children in Experiment 2, and in 6-year-olds in the combined analysis of Experiments 1 and 2. These findings imply that semantic coding may be used by some children, but the lack of consistency indicates that the presence of this type of coding is variable and difficult to detect.

The results of Experiment 3 also suggested that self-report measures of phonological coding in young children might be a reliable way of assessing these skills. Orderly effects of phonological similarity and word length were found in children who reported the use of phonological coding and these effects were absent in those who did not report the strategy. Of additional interest was the imperfect relationship between age and self-reported phonological coding, implying that the development of phonological coding may only be broadly related to chronological age.

In summary, three experiments addressed a number of controversies in the previous developmental literature on phonological and other forms of coding, using methodologically sound, sensitive and reliable assessments of picture memory span that controlled for potentially confounding effects of verbal input and output. We demonstrated that the emergence of verbal mediation in the form of phonological coding strategies is a robust and reliable stage of development in typical children, occurring from the age of five or six years. This confirmed findings initially reported by Conrad (1971), but subsequently questioned by other authors. However, in contrast to some previous research, little evidence was found for a distinct visual coding stage at 4 years, casting doubt on assumptions that this is a developmental stage that consistently precedes phonological coding. There was, however, some evidence for a dual visual and phonological coding stage prior to exclusive use of phonological coding at around 5-6 years. Evidence for semantic similarity effects was limited, suggesting that semantic coding is not a key method by which young children recall lists of pictures. Consequently, the findings provide new information about the fundamental ways in which children's representations and information processing change during the early primary school years.

References

Al-Namlah, A.S., Fernyhough, C. & Meins, E. (2006). Sociocultural influences on the development of verbal mediation: Private speech and phonological recoding in Saudi Arabian and British samples. *Developmental Psychology, 42*:1, 117-131. doi: 10.1037/0012-1649.42.1.117

Avons, S. E. (1998). Serial report and item recognition of novel visual patterns. *British Journal of Psychology*, *89*, 285–308. doi: 10.1111/j.2044-8295.1998.tb02685.x

Avons, S. E. & Mason, A. (1999). Effects of visual similarity on serial report and item recognition. *Quarterly Journal of Experimental Psychology*, *52A*, 217–240. doi: 10.1080/027249899391296

Avons, S. E., Ward, G. & Melling, L. (2004). Item and order memory for novel visual patterns assessed by two-choice recognition. *Quarterly Journal of Experimental Psychology*, *57A*, 865–891. doi: 10.1080/02724980343000521

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*, 362–365. doi: 10.1080/14640746608400055

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2007). *Working memory*, *thought*, *and action*. Oxford: Oxford University Press.

Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A. D., Thomson, N. & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589. doi: 10.1016/S0022-5371(75)80045-4

Baddeley, A.D., Lewis, V. & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology, 36A*, 233-252. doi: 10.1080/14640748408402157

Brown, R. M. (1977). An examination of visual and verbal coding processes in preschool children. *Child Development*, *48*, 38–45. doi: 10.2307/1128878

Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Psychology*, *5*, 398–405. doi: 10.1037/h0031595

Conrad, R. & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*, 429–432. doi: 10.1111/j.2044-8295.1964.tb00928.x

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T. & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, *34*, 1–17. doi: 10.1016/0749-596X(92)90002-F

Cowan, N., Saults, J. S., Winterowd, C. & Sherk, M. (1991). Enhancement of 4-year-old children's memory span for phonologically similar and dissimilar word lists. *Journal of Experimental Child Psychology*, *51*, 30–52. doi: 10.1016/0022-0965(91)90076-5

Ferguson, A.N., Bowey, J.A. & Tilley, A. (2002). The association between auditory memory span and speech rate in children from kindergarten to sixth grade. *Journal of Experimental Child Psychology, 81*, 141–156. doi: 10.1006/jecp.2001.2646

Flavell, J. H., Green, F. L., Flaevell, E. R. & Grossman: J. B. (1997). The development of children's knowledge about inner speech. *Child Development, 68*:1, 39-47. 10.1111/j.1467-8624.1997.tb01923.x

Ford, S. & Silber, K.P. (1994). Working memory in children: A developmental approach to the phonological coding of pictorial material. *British Journal of Developmental Psychology, 12*, 165-175. 10.1111/j.2044-835X.1994.tb00626.x

Frick, R. W. (1988a). Issues of representation and limited capacity in the visuospatial sketchpad. *British Journal of Psychology*, *79*, 289–308. doi: 10.1111/j.2044-8295.1988.tb02289.x

Frick, R. W. (1988b). Representation, recovery, and limited capacity in an auditory short-term store. *British Journal of Psychology*, *79*, 213–240. doi: 10.1111/j.2044-8295.1988.tb02284.x

Gathercole, S. E. & Hitch, G. J. (1993). Developmental changes in short-term memory: A revised working memory perspective. In A. Collins, S. E. Gathercole, M. A. Conway and P. E. Morris (Eds.), *Models of short-term memory*. Hove, UK: Psychology Press.

Halliday, M. S., Hitch, G. J., Lennon, B. & Pettipher, C. (1990). Verbal short-term memory in children: The role of the articulatory loop. *European Journal of Cognitive Psychology*, *2*, 23–38. doi: 10.1080/09541449008406195

Hasselhorn, M. & Grube, D. (2003). The phonological similarity effect on memory span in children: Does it depend on age, speech rate, and articulatory suppression? *International Journal of Behavioral Development*, *27*, 145–152. doi: 10.1080/01650250244000146

Hayes, D. S. & Schulze, S. A. (1977). Visual encoding in preschoolers' serial retention. *Child Development*, *48*, 1066–1070. doi: 10.2307/1128362

Henry, L. A. (1991). The effects of word length and phonemic similarity in young children's short-term memory. *Quarterly Journal of Experimental Psychology*, *43A*, 35–52. doi: 10.1080/14640749108400998

Henry, L. (2008). Short-term memory coding in children with intellectual disabilities. *American Journal of Mental Retardation, 113*, 187–200. doi: 10.1352/0895-8017(2008)113[187:SMCICW]2.0.CO;2

Henry, L. A. & Millar, S. (1993). Why does memory span improve with age? A review of the evidence for two current hypotheses. *European Journal of Cognitive Psychology*, *5*, 241-287. doi: 10.1080/09541449308520119

Henry, L. A., Turner, J. E., Smith, P. T. & Leather, C. (2000). Modality effects and the development of the word length effect in children. *Memory*, *8*, 1–17. doi: 10.1080/096582100387678

Hitch, G. J., Woodin, M. E. & Baker, S. (1989). Visual and phonological components of working memory in children. *Memory & Cognition*, *17*, 175–185. doi: 10.3758/BF03197067

Hitch, G.J, Halliday, M.S., Dodd, A. & Littler, J.E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology, 7*, 347–362. doi: 10.1016/0022-0965(91)90033-O

Hitch, G. J., Halliday, M. S., Schaafstal, A. M. & Heffernan, T. M. (1991). Speech, “inner speech” and the development of short-term memory: Effects of picture labeling on recall. *Journal of Experimental Child Psychology*, *51*, 220–234. doi: 10.1016/0022-0965(91)90033-O

Hitch, G. J., Halliday, S., Schaafstal, A. M. & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory & Cognition*, *16*, 120–132. doi: 10.3758/BF03213479

Hulme, C. (1987). The effects of acoustic similarity on memory in children: A comparison between visual and auditory presentation. *Applied Cognitive Psychology, 1*, 45-51. doi: 10.1002/acp.2350010107

Hulme, C., Thomson, N., Muir, C. & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology, 38*, 241–253. doi: 10.1016/0022-0965(84)90124-3

Hulme, C., Neath, I., Stuart, G., Shostak, L., Surprenant, A.M. & Brown, G.D.A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 586–594. doi: 10.1037/0278-7393.32.3.586

Huttenlocher, J. & Lui, F. (1979). The semantic organisation of some simple nouns and verbs *Journal of Verbal Learning and Verbal Behavior, 18*, 141–162. doi: 10.1016/S0022-5371(79)90091-4

Jarrold, C., Cocksey, J. & Dockerill, E. (2008). Phonological similarity and lexicality effects in children’s verbal short-term memory: Concerns about the interpretation of probed recall data. *Quarterly Journal of Experimental Psychology, 61* (2), 324-340. doi: 10.1080/17470210701202210

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, Mass.: MIT Press/Bradford Books.

Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, *22*, 1–38.

Lewandowsky, S. & Oberauer, K. (2008). The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin & Review, 15 (5),* 875-888. doi: 10.3758/PBR.15.5.875

Logie, R. H., Della Sala, S., Wynn, V. & Baddeley, A. D. (2000). Visual similarity effects in immediate serial recall. *Quarterly Journal of Experimental Psychology, 53A*, 626–646. doi: 10.1080/713755916

Longoni, A. M. & Scalisi, T. G. (1994). Developmental aspects of phonemic and visual similarity effects: Further evidence in Italian children. *International Journal of Behavioral Development*, *17*, 57–71. doi: 10.1177/016502549401700104

Marcovitch, S. & Zelazo, P.D. (2009). A hierarchical competing systems model of the emergence and early development of executive function. *Developmental Science, 12*(1)*,* 1–25. doi: 10.1111/j.1467-7687.2008.00754.x

McGilly, K. & Siegler, R.S. (1989). How children choose among serial recall strategies. *Child Development, 60*, 172-182. doi: 10.2307/1131083

Messer, D.J. & Dockrell, J. (2006). Children's naming and word-finding difficulties: Descriptions and explanations. *Journal of Speech, Language, and Hearing Research, 49*, 309-324. doi: 10.1044/1092-4388(2006/025)

Messer, D.J., Pine, K. J. & Butler, C. (2008). Children's behaviour and cognitions across different balance tasks. *Learning and Instruction, 18*, 42–53. doi: 10.1016/j.learninstruc.2006.09.008

Miller, P.H. (1994). Individual differences in children's strategic behaviour: utilisation deficiencies. *Learning and Individual Differences, 6*, 285-307.

Morrison, C. M., Chappell, T. D. & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, *50A*, 528–559. doi: 10.1080/027249897392017

Mueller, S.T., Seymour, T.L., Keiras, D.E. & Meyer, D.E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(6)*, 1353-1380. doi: 10.1037/0278-7393.29.6.1353

Nimmo, L. M. & Roodenrys, S. (2004). Investigating the phonological similarity effect: Syllable structure and the position of common phonemes. *Journal of Memory and Language*, *50*, 245–258. doi: 10.1016/j.jml.2003.11.001

Palmer, S. (2000a). Working memory: A developmental study of phonological recoding. *Memory*, *8*, 179–193. doi: 10.1080/096582100387597

Palmer, S. (2000b). Development of phonological recoding and literacy acquisition: a four-year course-sequential study. *British Journal of Developmental Psychology, 18*, 533–555. doi: 10.1348/026151000165841

Pine, K. & Messer, D. (2003). The development of representations as children learn about balancing. *British Journal of Developmental Psychology, 21:2,* 285–301. doi: 10.1348/026151003765264093

Purser, H. & Jarrold, C. (2010). Short- and long-term memory contributions to immediate serial recognition: Evidence from serial position effects. *Quarterly Journal of Experimental Psychology, 63*(4), 679-693. doi: 10.1080/17470210903067635

Romani, C., McAlpine, S., Olson, A., Tsouknida, E. & Martin, R. (2005). Length, lexicality, and articulatory suppression in immediate recall: Evidence against the articulatory loop. *Journal of Memory and Language*, *52*, 398–415. doi: 10.1016/j.jml.2005.01.005

Vygotsky, L.S. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, MA: MIT press. (Original work published 1934).

Whitehouse, A.J.O., Maybery, M.T. & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry, 47:*8, 857-865. doi: 10.1111/j.1469-7610.2006.01624.x

Williams, D. & Jarrold, C. (2010). Predicting inner speech use amongst children with autism spectrum disorder (ASD): The roles of verbal ability and cognitive profile. *Journal of Autism and Developmental Disorders*, *37*, 1413-1429 doi: 10.1007/s10803-010-0936-8

Williams, D., Happé, F. & Jarrold, C. (2008). Intact inner speech use in autism spectrum disorder: Evidence from a short-term memory task. *Journal of Child Psychology and Psychiatry, 49*:1, 51-58. doi: 10.1111/j.1469-7610.2007.01836.x

Winsler, A. & Naglieri, J. (2003). Overt and covert verbal problem-solving strategies: Developmental trends in use, awareness, and relations with task performance in children aged 5 to 17*. Child Development, 74*:3, 659-678. doi: 10.1111/1467-8624.00561

Yuzawa, M. (2001). Effects of word length on young children's memory performance. *Memory & Cognition*, *29*, 557–564. doi: 10.3758/BF03200457

Zelazo, P.D. (2004). The development of conscious control in childhood. *Trends in Cognitive Sciences, 8*, 12–17. doi: 10.1016/j.tics.2003.11.001

Table 1. Mean age of acquisition (AoA), imageability, frequency, familiarity and name agreement for each of the six sets of experimental picture span materials (from Morrison et al., 1997). Each line in italics gives the range of values for the preceding set. Lower AoA values refer earlier rated age of acquiring the word; for imageability, frequency and familiarity, higher values indicate higher ratings; perfect name agreement is given by a value of 1.00.

Picture Type	AoA (scale 1-7)	Imageability (scale 1-7)	Frequency (scale 1-5)	Familiarity (scale 1-5)	Name agreement (scale 0-1)
Control	2.01	6.37	3.17	3.38	0.96
(Range)	*1.75-2.50*	*5.95-6.70*	*2.00-4.15*	*2.09-4.77*	*0.82-1.00*
Phonological	2.23	6.24	3.04	3.50	0.95
(Range)	*1.15-2.90*	*5.90-6.70*	*2.35-3.90*	*2.59-4.70*	*0.86-1.00*
Visual	2.03	6.29	3.77	4.03	0.96
(Range)	*1.25-2.45*	*5.80-6.75*	*2.35-4.70*	*2.82-4.86*	*0.82-1.00*
Long	2.11	6.49	3.11	3.51	0.84
(Range)	*1.70-2.45*	*6.25-6.70*	*1.90-4.35*	*2.20-4.59*	*0.59-1.00*
Semantic	1.64	6.16	3.50	4.39	0.93
(Range)	*1.45-2.00*	*5.75-6.70*	*2.85-4.25*	*2.88-4.73*	*0.68-1.00*
Control 2	2.01	6.41	3.52	3.94	0.98
(Range)	*1.30-2.55*	*6.00-6.80*	*2.50-4.40*	*2.50-4.68*	*0.91-1.00*

Table 2: Mean memory spans (Experiment 1), mean trials correct (Experiment 2) and mean proportions correct (Experiment 3) for control, phonologically similar, visually similar, long-named and semantically similar picture items: 4/5-year olds and 6/7/8-year-olds (SD in brackets).

	Picture Span Item Type				
	Control	Phonologically Similar	Visually Similar	Long-named	Semantically Similar
			Experiment 1 (Span Scores)		
4/5-year-olds	1.67 (0.84)	1.39 (0.61)	1.50 (0.71)	1.33 (0.49)	1.61 (0.70)
6/8-year-olds	3.19 (1.11)	2.30 (0.87)	3.22 (1.15)	2.52 (0.85)	3.11 (0.97)
			Experiment 2 (Trials Correct)		
4/5-year-olds	5.81 (2.17)	5.08 (1.38)	5.12 (1.88)	5.23 (1.80)	5.31 (1.52)
6/7-year-olds	8.43 (2.87)	6.30 (1.75)	7.00 (1.89)	6.70 (1.49)	7.73 (2.38)
			Experiment 3 (Proportion Correct)		
4/5-year-olds	.44 (.24)	.40 (.20)	.38 (.20)	.42 (.23)	.44 (.22)
6/8-year-olds	.53 (.25)	.44 (.23)	.56 (.27)	.44 (.19)	.53 (.23)

Captions:

Figure 1: Mean span scores on control, phonologically similar, visually similar, long-named and semantically similar pictures for high, intermediate and low performing clusters in Experiment 1 (with standard error bars).

Figure 2: Mean trials correct on control, phonologically similar, visually similar, long-named and semantically similar pictures for high, intermediate and low performing clusters in Experiment 2 (with standard error bars).

Figure 3: Mean proportion correct on control, phonologically similar, visually similar, long-named and semantically similar pictures for high and low performing clusters in Experiment 3 (with standard error bars).

Figure 4: Mean span scores on control, phonologically similar, visually similar, long-named and semantically similar pictures for children of ages 4, 5, 6 and 7/8 in Experiment 1 and Experiment 2 combined (with standard error bars).

Figure 1:

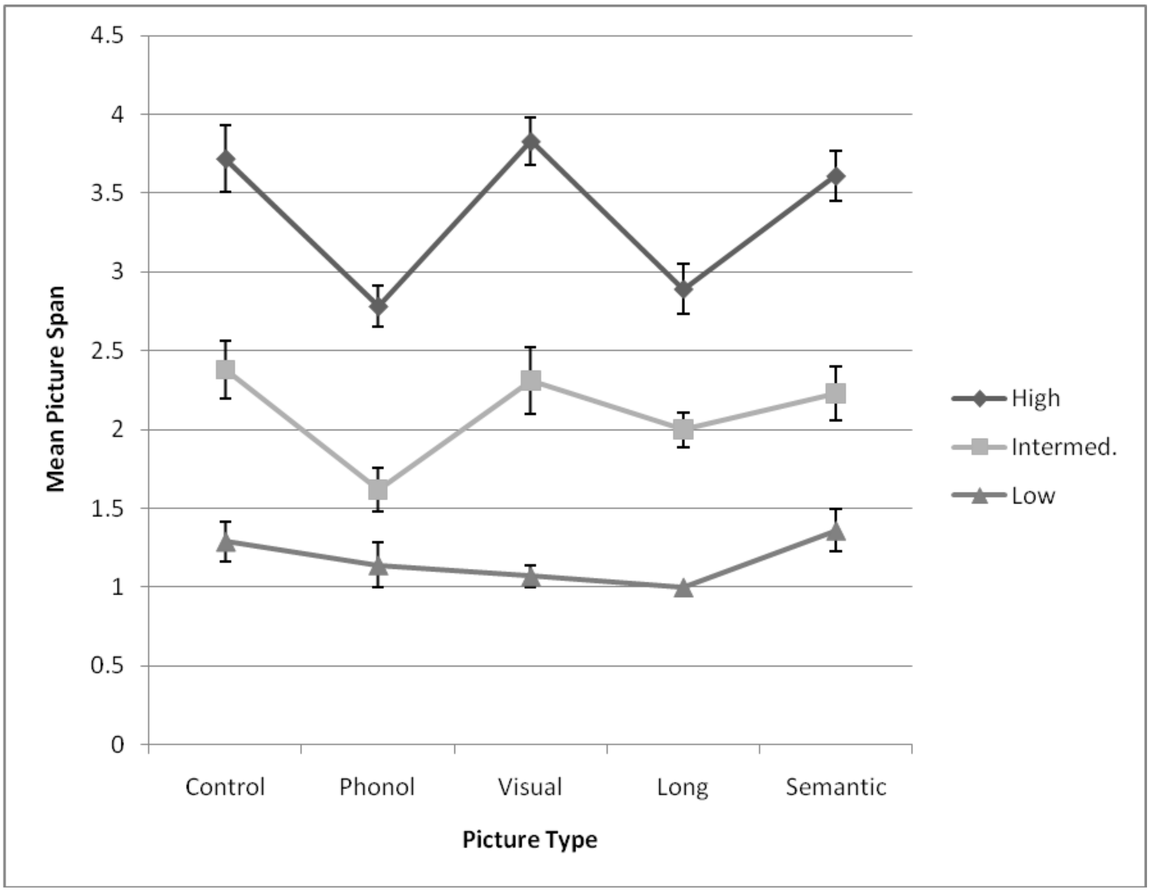
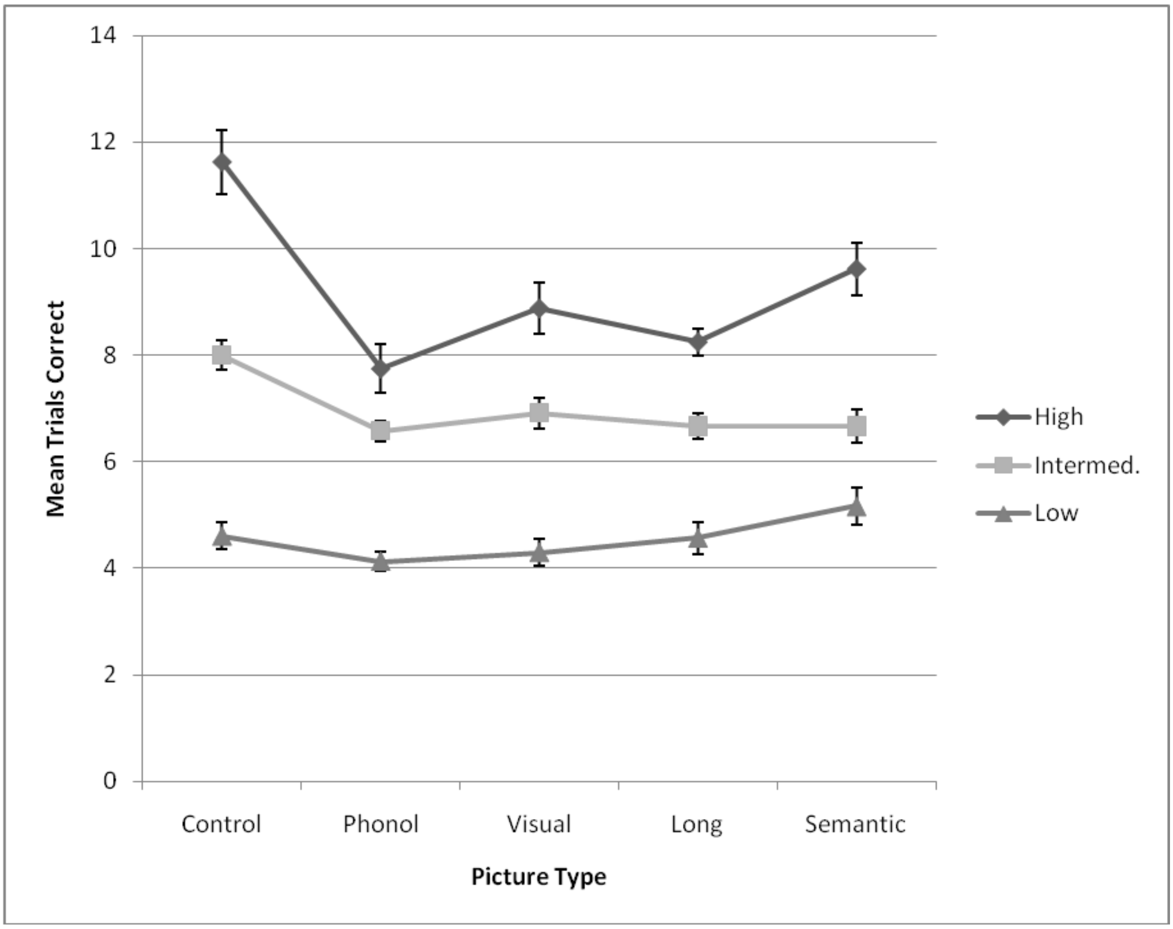

4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Mean Picture Span
Control
Phonol
Visual
Long
Semantic
Picture Type
High
Intermed.
Low

Figure 2:



14
12
10
8
6
4
2
0
Mean Trials Correct
Control
Phonol
Visual
Long
Semantic
Picture Type
High
Intermed.
Low

Figure 3:

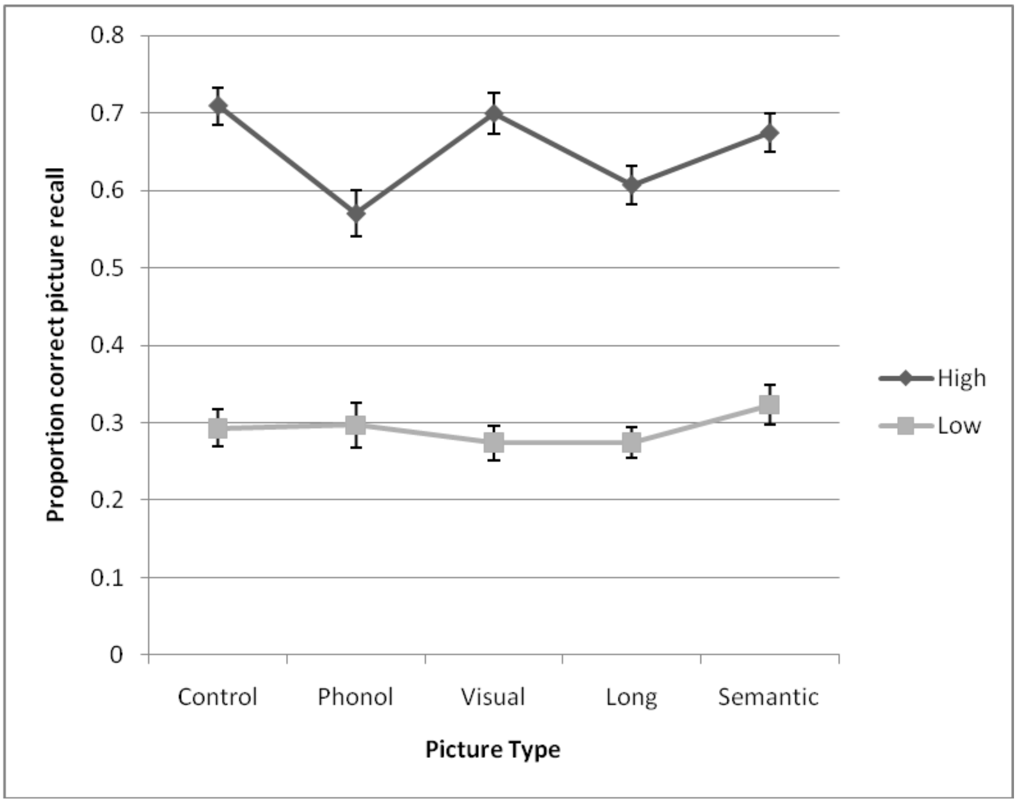
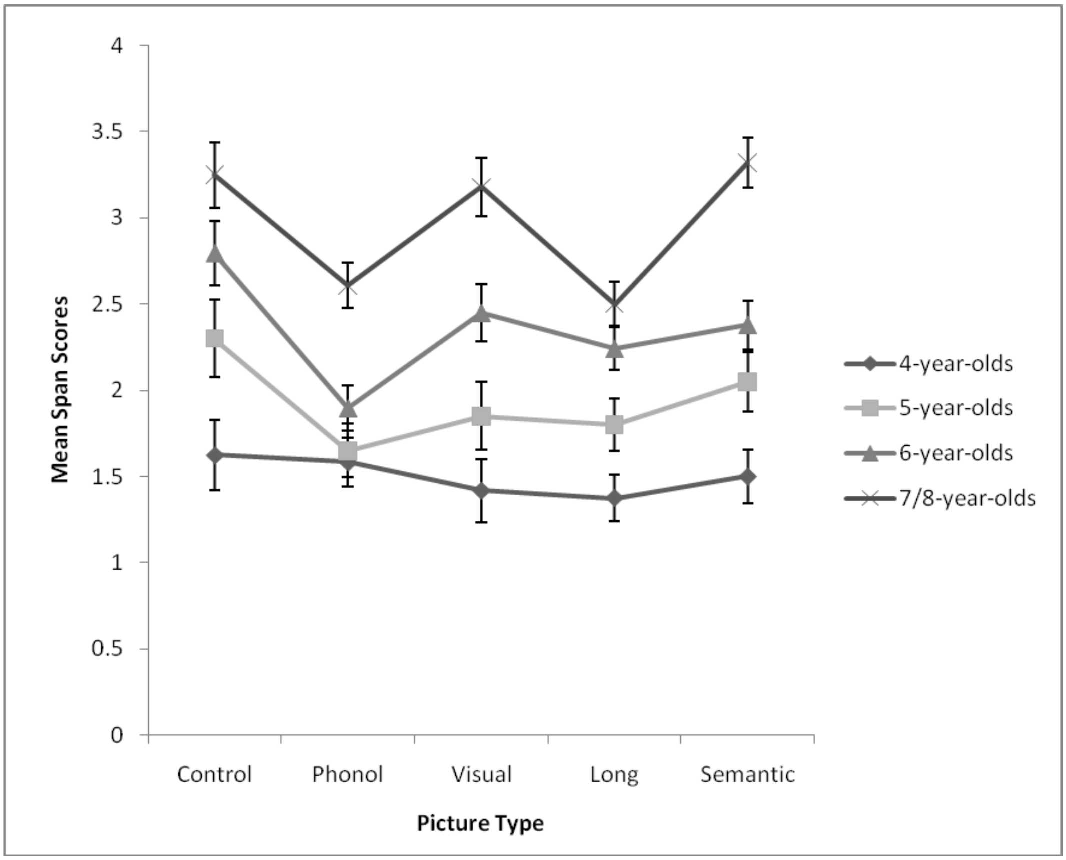

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Proportion correct picture recall
Control
Phonol
Visual
Long
Semantic
Picture Type
High
Low

Figure 4:


4
3.5
3
2.5
2
1.5
1
0.5
0
Mean Span Scores
Control
Phonol
Visual
Long
Semantic
Picture Type
4-year-olds
5-year-olds
6-year-olds
7/8-year-olds

Appendix. **Questions about phonological coding**

1. **Can you tell me a bit about how you tried to remember the pictures when I showed them to you?** If the child gives an answer about pictures/words or some other strategy then ask –

2. **How often did you do this? – most of the time, just sometimes, or hardly at all?** Ask these follow up question *unless* the child has already very clearly described their strategy:

3. **Some children just look at the pictures on the cards and then hope they will remember which pictures they have seen when they are shown the big card did you do this? If yes – was it most of the time, just sometimes, or hardly at all?**

4. **Other children try to remember by thinking about what they have seen in their head, to help them remember they try to think about the picture they have seen and what the pictures looked like – did you do this? If yes – was it most of the time, just sometimes, or hardly at all?**

5. **There also are children who try to remember the pictures by saying the name of the picture in their head – did you ever try this? If yes – was it most of the time, just sometimes, or hardly at all?**